***Relatório - Andamento da Rede Neural Geradora de Road Maps***

*05/09/2023*

*Objetivos Semanais*

| **03 a 09/09** | **10 a 16/09** | **17 a 23/09** | **24 a 30/09** |
| --- | --- | --- | --- |
| Negociar computador do IFES para treinamento. | Buscar máquinas virtuais para treinamento da rede. |  |  |
| Realizar treinamento prévio (curto) com Checkpoints e rede custom. | Pré processar novos dados de treinamento e teste |  |  |
| Documentar sobre os Checkpoints. | Documentar sobre a rede UNET custom da biblioteca UNET. |  |  |
| Verificar a existência de outras técnicas para otimizar o treinamento. | Documentar sobre o processo de pré-processamento de dados |  |  |
| Documentar sobre a escolha do Kaggle pelo Colab. | Escrever sobre as barreiras da implementação transfer learning brevemente |  |  |

O ambiente virtual escolhido para se executar o treinamento da Rede Neural tem um limite de 12 horas seguidas de uso de sessão de kernel, após esse limite a sessão para e é necessário que o usuário a reative. Sendo assim, na tentativa de se realizar um treinamento utilizando o Kaggle, mesmo que curto, procurou-se aplicar técnicas para otimização de tempo e salvamento do treinamento.

* **ModelCheckpoint [1]**

Essa ferramenta faz parte da biblioteca Keras e nos permite salvar o ponto em que o treinamento foi finalizado ou interrompido. Sendo assim, é possível recomeçar o treinamento do último checkpoint salvo. Outro bom uso do checkpointing é gerar os pesos do modelo cada vez que uma melhoria é observada durante o treinamento, assim os melhores valores do treinamento são salvos.

* **EarlyStopping [2]**

Uma abordagem também muito interessante é o EarlyStopping, um tipo de retorno de chamada no Keras que encerra o treinamento quando uma métrica monitorada para de melhorar. É uma forma de evitar o overfitting, interrompendo o processo de treinamento quando o modelo começa a aprender demais o conjunto de dados de treinamento [3].

* **Label Smoothing [6]**

Ao aplicar a suavização de rótulos, podemos diminuir a confiança do modelo e evitar que ele caia em fendas profundas do cenário de perdas onde ocorre o sobreajuste. Evitando assim que o modelo fique demasiado confiantenas suas previsões. Entre os métodos de Label Smoothing temos: a suavização de rótulos atualizando explicitamente sua lista de rótulos e suavização de rótulo usando a função de perda do Keras. É recomendado o uso do Label Smoothing quando se estiver tendo problemas para generalizar o modelo e/ou seu modelo estiver se ajustando demais ao seu conjunto de treinamento.

* **Mixed Precision [7]**

Podemos definir este tipo de treinamento para fazer um modelo aprender pontos flutuantes de 16 e 32 bits para que o modelo possa aprender mais rápido e o tempo de treinamento do modelo possa ser reduzido [4]. A precisão mista é o uso de tipos de ponto flutuante de 16 bits e 32 bits em um modelo durante o treinamento para torná-lo mais rápido e usar menos memória. Ao manter certas partes do modelo nos tipos de 32 bits para estabilidade numérica, o modelo terá um tempo de etapa menor e treinará igualmente em termos de métricas de avaliação, como precisão.

* **Kaggle [9] x Colab [8]**

Diante das limitações de hardware existentes, optou-se por utilizar-se um ambiente que contenha máquina virtual, porém as plataformas disponíveis de forma gratuita ou acessível contém recursos limitados também. Sendo assim, visou se analisar entre as principais ferramentas disponíveis e mais utilizadas: Google Colab (que possui plano gratuito e pro) e Kaggle.

|  | **Kaggle** | **Google Colab Free** | **Google Colab Pro** |
| --- | --- | --- | --- |
| **Vantagens** | * Integração com a criação e armazenamento de datasets do próprio Kaggle * Mais variedade de GPU | * Memória Ligada ao Google Drive * GPU sem limites de horas semanais | * Memória Ligada ao Google Drive * Variedade de Hardware * Hardwares Melhores * 24 horas de sessão continua |
| **Desvantagens** | * 12 horas de sessão continua. * 30 horas de GPU semanais * Memória de output limitada | * Uma única opção de GPU * 12 de sessão continua * GPU pode não estar disponível | * Uso dos serviços limitados a créditos * Se houver muitas pessoas no google cloud não haverá GPUs disponíveis. * Recurso Pago |

**Sugestões da Rachel:**

[**https://gitlab.inf.unibz.it/Rachel.FantiCoelhoLima/intelligent-lane-detection-in-low-cost-devices/-/blob/main/model\_ultra\_variants/ultra\_tf\_grid/train\_ultra\_tf.ipynb?ref\_type=heads**](https://gitlab.inf.unibz.it/Rachel.FantiCoelhoLima/intelligent-lane-detection-in-low-cost-devices/-/blob/main/model_ultra_variants/ultra_tf_grid/train_ultra_tf.ipynb?ref_type=heads)

* **tf.py\_function para ler arquivos**
* **usar lane\_train\_ds.map(lane\_ds.process\_json, num\_parallel\_calls=PARALLEL\_JOBS)**
* **grid ao invés de pegar todos os pontos da imagem**
* **Diminuir tamanho da imagem**
* **De todas, as variáveis de uma rede neural, as que são muito importantes de serem bem parametrizadas no modelo de ML são: Learning rate, o tamanho dos batches e o número de hidden layers. Essas impactam significativamente na performance do modelo.De todas, as variáveis de uma rede neural, as que são muito importantes de serem bem parametrizadas no modelo de ML sao: Learning rate, o tamanho dos batches e o número de hidden layers. Essas impactam significativamente na performance do modelo.**

|  | **Kaggle** | **Amazon**  **(AWS)** | **Azure** | **Google Colab** | **Google Colab Pro** | **Colab**  **Pro+** | **Google Cloud** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Cobrança** | Gratuito | Preço mensal varia de acordo com as configurações da VM escolhida  Possibilidade de diminuir o valor limitando a quantidade de horas mensais que será utilizada a VM. Descontos para planos anuais, pré pagos. | O valor pago mensalmente varia de acordo com as configurações da VM escolhida.  Planos anuais possuem desconto de 17% que podem ser pagos por mês. | Gratuito | Planos mensais | Planos mensais | Preço mensal varia de acordo com as configurações da VM escolhida |
| **Uso de recursos** | 30 horas semanais de GPU | Recursos pré pagos | Recursos pré pagos | Uso de GPU depende da disponibilidade. | Uso de créditos(units) acumulativos porém vencem em 90 dias. 100 units por mês. | Uso de créditos(units) acumulativos porém vencem em 90 dias. 500 units por mês. | Recursos pré pagos |
| **Tempo de sessão** | 12 horas seguidas de uso de sessão | - | - | 12 horas seguidas de uso de sessão | 24 horas de sessão continua | Dependendo da quantidade de créditos, ela é gasta por hora de acordo com o uso e com a categoria. Mas sua execução pode ultrapassar 24 horas contínuas.  Possibilidade da sessão ser executada até mesmo com navegador fechado. | 730 horas por mês. |
| **Hardwares disponíveis** | x2 GPU T4  GPU P100  TPU VM v3-8 | - g4dn.xlarge  4 vCPUs  16 GB de memória  Armazenamento de 150GB SSD  Performance acima de 10 Gigabit  US$0.37853/hora  - g4ad.xlarge  4 vCPUs  16 GB de memória  Armazenamento de 150GB SSD  Performance acima de 25 Gigabit  US$0.526/hora  -g4dn.2xlarge  8 vCPUs  32 GB de memória  Armazenamento de 225GB SSD  Performance acima de 25 Gigabit  US$0.752/hora | - NV6ads A10 v5; 6 vCPUs, 55GB de RAM, 180 GB de armazenamento temporário.  (US$ 0,454/hora)  - NV12ads A10 v5; 12 vCPUs, 110GB de RAM, 320 GB de armazenamento temporário.  (US$ 0,908/hora)  - NC4as T4 v3: 4 vCPUs, 28GB de RAM, 180GB armazenamento temporário, US$ 0,526/hora  - NC8as T4 v3: 8 vCPUs, 56GB de RAM, 360GB armazenamento temporário, US$ 0,752/hora  - NC6, 6 vCPUs, 56GB de RAM, 340GB de armazenamento temporário, US$ 0,9/hora | T4 GPU, TPU | GPU Standard e Premium | GPU Standard e Premium | - a2-highgpu-1g  NVIDIA TESLA A100  12 vCPUs  85GB de RAM  US$2,681.57/mês  - n1-standard-8  1 NVIDIA TESLA K80  8 vCPUs  30GB de RAM  US$605.90/mês  - n1-standard-8  1 NVIDIA TESLA P100  8 vCPUs  30GB de RAM  US$1,343.20/mês  - n1-standard-4  1 thread por core  1 NVIDIA TESLA K80  8 vCPUs  30GB de RAM  US$467.20/mês |
| **Configurações** | RAM: 13GB  GPU Mem: 15.9GB  DISK: 73.1GB  Arquivos de Saída: 19.5GB | Depende do item acima | Depende do item acima | RAM: 12.7GB  DISK: 107.7GB | RAM: 25GB  DISK: aprox. 200GB | RAM: 52GB  DISK: aprox. 200 GB | RAM depende do item acima.  DISCO:  - SSD 128GB  USD 21.76  - SSD 256GB  USD 43.52 |
| **Média de valores** | - | US$0,552/hora  US$201,48/mês (considerando um uso de 50% das horas de um mês) | US$0,708/hora  US$258,42/mês (considerando um uso de 50% das horas de um mês) | - | R$48,00  /mês | R$258,00  /mês | US$1,745/hora  US$1270,00/mês  (considerando 100% das horas)  \*Sem considerar disco de armazenamento\* |
| **Observações** | - | 750 horas gratuito para uso do t2 micro\* por 12 meses | 200 dólares de crédito | - | - T4 GPU gasta cerca de 1.96 units por hora;  - A100 gasta gasta cerca de 13.08 units por hora | - T4 GPU gasta cerca de 1.96 units por hora;  - A100 gasta gasta cerca de 13.08 units por hora | 300 dólares de crédito válidos por 12 meses |

**T2 Micro\***

**1 vCPU, 1GB Memória, performance baixa a moderada.**

**Calculadora Azure:**

[**https://azure.microsoft.com/pt-br/pricing/calculator/**](https://azure.microsoft.com/pt-br/pricing/calculator/)

**Calculadora AWS:**

[**https://calculator.aws/#/addService/ec2-enhancement**](https://calculator.aws/#/addService/ec2-enhancement)

**Calculadora Google:**

[**https://cloud.google.com/products/calculator#id=**](https://cloud.google.com/products/calculator#id=)

* **Desafios do Transfer Learning**

*Situação*

| **03 a 09/09** | **10 a 16/09** | **17 a 23/09** | **24 a 30/09** |
| --- | --- | --- | --- |
| Negociar computador do IFES para treinamento. | Buscar máquinas virtuais para treinamento da rede e escrever sobre. | Treinamentos com diferentes valores de parâmetros (dados 10k). |  |
| Realizar treinamento prévio (curto) com Checkpoints e rede custom. | Pré processar novos dados de treinamento e teste |  |  |
| Documentar sobre os Checkpoints. | Documentar sobre a rede UNET custom da biblioteca UNET. |  |  |
| Verificar a existência de outras técnicas para otimizar o treinamento. | Documentar sobre o processo de pré-processamento de dados |  |  |
| Comparar plataformas Kaggle e Colab. | Verificar sugestões e funções propostas pela Rachel no modelo de treinamento. | Escrever sobre as barreiras da implementação transfer learning brevemente |  |
|  | Treinamento com os novos dados gerados (10K) |  |  |

***Legenda:***

*Vermelho - Não iniciado*

*Verde - Finalizado*

*Amarelo - Iniciado mas não finalizado*

*Cinza - Alterado/Cancelado*

**Referências Bibliográficas**

[1] TEAM, K. Keras documentation: ModelCheckpoint. Disponível em: <<https://keras.io/api/callbacks/model_checkpoint/>>.

[2] TEAM, K. Keras documentation: EarlyStopping. Disponível em: <https://keras.io/api/callbacks/early\_stopping/>.

[3] Understanding the Mystery of EarlyStopping Callback in Keras | Saturn Cloud Blog. Disponível em: <https://saturncloud.io/blog/understanding-the-mystery-of-earlystopping-callback-in-keras/#:~:text=EarlyStopping%20is%20a%20type%20of>.

‌[4] VERMA, Y. 7 tricks to speed up the training of a neural network. Disponível em: <https://analyticsindiamag.com/7-tricks-to-speed-up-the-training-of-a-neural-network/>.

‌[5] KIRCH, S. Speed up your Training with Mixed Precision on GPUs and TPUs in TensorFlow. Disponível em: <<https://towardsdatascience.com/speed-up-your-tensorflow-training-with-mixed-precision-on-gpu-tpu-acf4c8c0931c>>.

[6] ROSEBROCK, A. Label smoothing with Keras, TensorFlow, and Deep Learning. Disponível em: <https://pyimagesearch.com/2019/12/30/label-smoothing-with-keras-tensorflow-and-deep-learning/>.

[7] Precisão mista | TensorFlow Core. Disponível em: <https://www.tensorflow.org/guide/mixed\_precision?hl=pt-br>. Acesso em: 6 set. 2023.

[8] GOOGLE. Google Colaboratory. Disponível em: <https://colab.research.google.com/>.

[9] KAGGLE. Kaggle: Your Home for Data Science. Disponível em: <https://www.kaggle.com/>.

‌

‌

‌

‌